

マルチメディア研究開発の事例紹介

藤本 眞+ 浅原重夫++ 前川 肇+++ 浜田正宏+ 中川雅通+

松下電器産業株式会社

中央研究所+ 映像音響情報研究所++ オーディオビデオ研究所+++

コンピュータと通信技術の進展によって、メディアはより「納得できる」あるいは「楽しい」といった感性に踏み込む処理技術をもってさらに進展しつつある。従来、このような次世代を目指してきた技術が徐々に現実のものとして我々の目前に現れつつある状況を事例を通して概観する。

A Review of Multi Media Systems and Technologies by Research Examples

Makoto Fujimoto+ Shigeo Asahara++ Hajime Maekawa+++
Masahiro Hamada+ Masamichi Nakagawa+

Central Research Labs.+ Video Audio Information Technology Lab.++
Audio and Video Research Lab.+++
Matsushita Electric Industrial Co., Ltd.

Recent progress of computer and communication technologies have enabled Media application that are useful and enjoyable for everybody. This is being accelerated by new technologies that give us deeper understanding and more enjoyment with communication Media. Some examples of Multi-Media technologies are outlined here. They demonstrate some research trends with the coming Media.

1. はじめに

コンピュータやネットワークが提供する高度で多様な情報処理や通信サービスが家庭に影響を与え始めている。その一方、家庭における情報はオフィスのそれと用途や利用方法に大きな隔たりがあり、直接的な技術導入は判り難い、使いづらいという形で普及を阻む要因になっている。

ところで、家庭における情報化の真の目的は利用者の意図を理解した新しいサービス、今までにない体験の提供にある。このためには音声・画像・言語・図形・音響などのメディアを介した機械との自然で快適な対話をおこなう情報処理機能やインターフェイスの構築が必須となる。

ここでは次世代映像音響情報技術の展望を踏まえてマルチメディア技術研究の事例を紹介する。

2. ハードとソフトの進展とメディアの進化

情報処理を行うコンピュータの進展と情報を伝搬する通信技術の進化がハード面での基本となる。その技術水準は基盤となるの半導体と情報記録技術から推測される。光ディスク記録密度の進歩を2000年には10 bit/ μm^2 に達することが予想される。これは、1 Gbitを10 cm^2 に納める、すなわち映像、音響、文字、情報などを十分に蓄積できる記憶媒体の実現を示している。次に、この大量の情報を処理するコンピュータは、ダウングレードのトレンドにあり、同様に2000年には、マイクロプロセッサが予測される。

このように、映像・音響情報を手軽に処理する、あるいは、一般の電子機器に導入される技術基盤が整いつつある。

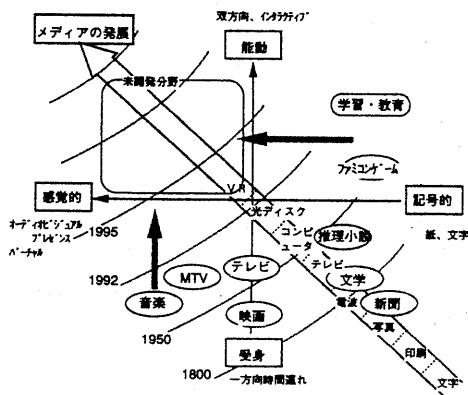


図1 メディアの進展

上述したコンピュータ並びに通信技術の高速化、高密度化は電子機器に高度にインテリジェントな機能を実現する、あるいはしつつある。こため旧来情報を一方的に提供するパッシブな存在であった電子機器が、状況に応じた対応ができるアクティブ性を持つことが可能になってきた。すなわち、受身な情報源であったテレビ、ステレオ、ビデオ機器が、アクティブでインタラクティブな性格を持ちつつある。これらの上に提供されるソフトウェアは図3のように進化の過程をたどると思われる。

図3の横軸は、記号的対視聴覚などの感覚的な軸を示す。横軸の先には3次元やVRなどの超臨場感の世界がある。一方、縦軸はポジティブ対アクティブの軸であり、旧来の活字、映画、テレビに代表される受身のメディアからTVゲームに代表される参加型メディアへの発展を示す。この延長にインタラクティブ映画、双方向テレビなどが考えられる。

ソフトウェア側から見たマルチメディアの進化の先は、この座標系における左上の象限、すなわちオーディオビジュアルかつアクティブの部分に存在する。これらが双方向テレビ、立体ゲーム、インタラクティブビデオなどの形となって現れると考えられる。このようにして新しく生まれると思われる技術は知的にも感覚的にも、よりインタラクティブで実在感をともなう超臨場感とも言う特徴をもつと考えられる。

以下にこれを支える技術研究の事例としてハード、ソフト、およびシステムの観点から、リアルタイム映像生成システム、モデルベース顔画像合成とマルチメディア映像編集装置およびマルチメディア会議システムを紹介する。

3. リアルタイム画像生成システム

コンピュータグラフィック技術はCAD/CAM/CAE、映像技術、ビジュアルシミュレーションなど多岐にわたり、より高速に高画質の映像を描画する機能が求められている。さらに、表示画像と外部イベントを同期させ、リアルタイム応答特性を生かして新しい応用形態を切り開くリアルタイム画像生成への要求も高まっている。これの事例として、画像生成に特化したアーキテクチャを有し、CGによる仮想空間内で物体間の相互作用を計算しながらリアルタイムに現実感の高い画像を生成するグラフィックシステムAVIP (Audio Video Information Processor)[1][2]を紹介する。

自然な画像を生成するにはレイトレーシングやラジオシティと呼ばれる計算処理を高速に実行する必要がある。AVIPは、高速バスやトータスネットワークと呼ぶアーキテクチャを用いた並列演算装置とこれを有効に生かす処理アルゴリズムおよびオペレーションシステムによって実現されている。

AVIPの全体構成を図2に示す。AVIPは疎結合の並列計算機であり、その単位となる計算ノードに対応するプロセッサユニットはそれぞれが一枚ボードであり、複数のRISCプロセッサとグラフィックスハードウェア、通信プロセッサから構成される。

RISCプロセッサは生成画像に依存する物体間の相互作用をシュミレーションするプログラムを実行しながらその結果を反映してリアルタイムに画像生成する。プロセッサユニットで生成された画像データはZマージ処理を行うZマージャと画像データバス(50MPixel/s)により高速にフレームメモリに転送される。画像生成に必要なデータはI/Oプロセッサに接続されたディスクやホストのワークステーションから各プロセッサユニットにトータスネットワークとバックプレーンバスを通じて入力される。10枚のプロセッサユニットは1ラック内に収納される。ラック内ではバックプレーンバスが使用でき、ラック間では大量のデータのバースト転送を行う通信路(200MB/s)が用意されている。

サンプリング法による並列化アルゴリズム[3]はAVIPの性能を十分引き出すことが可能である。

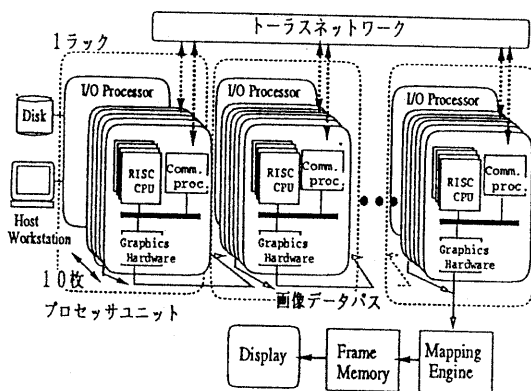


図2 AVIPの構成

ノード数20程度の場合ほぼ理想的に20倍の性能が望める。このシステムで20万パッチのデータを処理させると直接光ラジオシティ計算が光源毎に10秒弱で計算でき、パッチエレメントを基本とする仮説光ラジオシティ計算も1ループ毎に10秒弱で計算できるので、ループを100回実行しても16分半程度で計算できることになる。

このようにAVIPでは、100のユニット構成時のピーク特性はテクスチャ・照明マッピング、アンチエイアス付き独立三角形のHDTV(NTSC)CG画像を生成する場合、1200ポリゴン/秒に達し、リアルタイムな映像生成を目指したシステムとなっている。

4. モデルベース顔画像合成

コンピュータによる画像の合成が、インタラクティブな新しいメディアの世界を開けることは、すでに行われているシミュレーションによる可視化やエンターテイメントにおけるゲームにも見られる。このような技術はより自然な画像へと向かいと考えられる。この動向に対応してモデルベース技術とこれによる顔画像の合成技術システムを考える。

モデルベース符号化[4]は、画像に関するルール(知識)を用いて画像を認識、合成する手法ある。図3は、手法を用いた画像合成システムの基本構成[5]である。まず顔など物体の表面を小さなポリゴンの集まりで表した3次元モデルを生成し、その上に物体の2次元画像をテクスチャマッピング(貼り付け)を行って、画像を合成する。この手法では、物体の3次元構造モデルをルールに基づいて変化させることができ、これによって表情を持った顔など、色々な画像を合成することが可能である。このようにして物体の構造モデルを基に、一枚の顔写真から3次元的な動きや表情変化を行わことができる。

ここに示すシステムでは、これを並列処理可能なメディア処理ボードを用いて、パーソナルコンピュータ上で構成した。さらに入力画像から顔の特徴点を抽出することにより、3次元構造モデルを、その場で入力した顔画像から自動生成している。

各顔画像に合った3次元構造モデルの自動生成は、まず輪郭、目、鼻、口などの特徴点を濃淡情報、色相情報などを用いて抽出し、次に、それら特徴点を手がかりに、標準の3次元構造モデルの各多角形をなめらかに接続し、各3次元構造モデルを生成する。また、画像合成は以下のルールを用いている。

- 1) 動きルール : 顔の向き、口や目の動き、眉や頬の変化の仕方のルール
- 2) 表情ルール : 表情は、顔の動きルールの組み合わせ。その組み合わせ方をルール
- 3) 年齢ルール : 年齢変化は、顔の動きや表情変化のほかに骨格の変化を用い、成長する頭蓋の形状の変化、老年齢化によるたるみを表す。
 - ・幼年への変化ルール : 心理学の分野で頭蓋骨を横から見た形を逆ハート形のカーゴイド曲線で表現する。子供の頭蓋骨の成長を、このカーゴイド曲線のパラメータの変化するルール。
 - ・老年齢への変化ルール : 目の両端、鼻の付け根などの皮膚と骨の結合が強い部分を固定点として、皮膚の下方向へのたるみを放射線と正規分布関数で近似た、皮膚のたるみ、特に目尻、頬のたるみのルール。

これにより年齢変化を含むさまざまな表情を合成できるシステムとなっている。図4は、表情変化の例である。

この技術は、美容院や整形外科のシミュレーション、犯罪操作のためのモンタージュシステムなどへの幅広い応用が考えられる。また、音声に同期して口を動かし一枚の写真を喋らせ、コンピュータが表情をもって説明を行うプレゼンテーションシステムなどへの応用が考えられる。将来のマルチメディア通信やコンピュータによるプレゼンテーション、娯楽などの分野で柔軟に変化できるシステムが期待される。

5. マルチメディア指向映像編集機器

パソコン上でも動画像が扱えるようになったために、ビデオ映像の編集装置や音声・映像を有機的につなぐマルチメディアソフトの画像編集の研究がなされ始めている。従来の映像編集システムは専門の利用者を想定し、タイムコード等のキャラクターを用いたインターフェースを提供していた。このため、一般ユーザにとって分かりやすい、使い易いとは言えないものである。

ビデオの編集は、素材となる映像をそれを構成する映像の断片に分解し、それら映像の断片を編集の目的に従って検索・連結して再生することを繰り返す作業である。ところで、ビデオ編集の方法としては、VTRによる複写を中心とした編集方法が知られている。一方、映像の編集方法には旧来映画の編集に用いられてきた、フィルムを切り貼りして組み合わせるフィルムスタイルと呼ばれる方法があ

り、現在でも映画製作に用いられている。このフィルムスタイルは

- 1) 記録された映像を直接見て確認しながら編集作業ができる。
- 2) 編集途中で作られるフィルムの断片を組み合わせる編集作業ができ、微妙な調節であるトリミングや構成の変更が容易である。

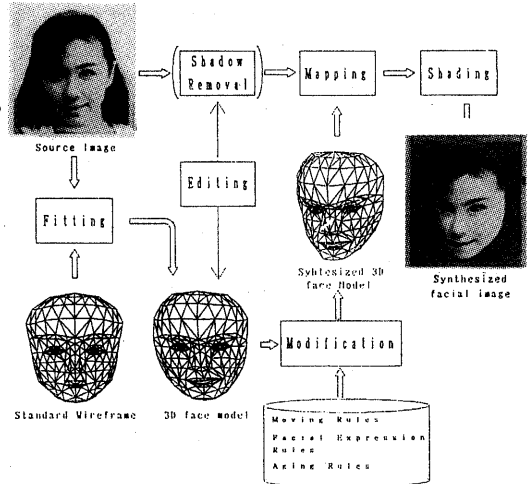


図3 モデルベース合成

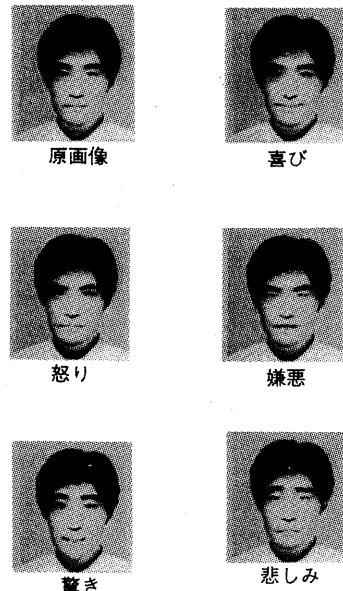


図4 表情変化

3) 編集を完了したフィルムを直接に再編集することが容易にできる。

という点で、通常のビデオ編集に比べて優れた方法となっている。これは映像の持つ時間的・空間的な情報の構造が視覚化されており、しかもそれを直接に操作できるので、利用者にとって分かりやすいためである。コンピュータ上での編集技術はこのような利点をビデオ編集に再現しようとするものである。

図5は、このような技術でワークステーションに構成された編集システム一例である[6]。このシステムは、以下の3つのソフトウェアで構成している。

- (1) Scene Maker - 必要な動画を探し出し、断片(セグメント)に切分ける。
- (2) Scene Icon - 切り出したセグメントを静止画アイコンとして管理する。
- (3) Scene Assembler - シーンアイコンの管理と、セグメントを望む順序に並べ変えて編集を行う。

また、本編集システムは機器制御型であり、ソースはAV機器上に存在する。そのため動画を表示し、音を鳴らすためにはソースを供給するAV機器を制御する必要がある。DCS(Divice Control System)は、AV機器を集中制御するための、サーバ・クライアント・モデルのシステムである。

このようにリモコンやフィルムのメタファの用いたインタフェイスで、フレームナンバー等のビデオ編集特有の複雑さを隠蔽、編集の基本要素を分り易くしている。このように、ビデオ編集に関する知識のないユーザにも簡単に操作ができ、学習しやすいシステムが実現されつつある。

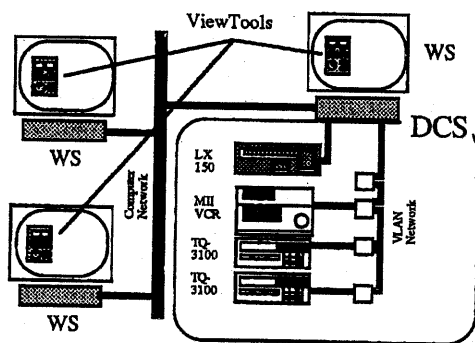


図5 システム構成

6. マルチメディア会議システム

コンピュータによる会議システムは、オフィスにおけるマルチメディア化の代表的な事例であり、すでに数多くが開発されている[7]。これには会議の臨場感向上を図るアプローチと会議内容の情報を利用して議事内容の理解を助るアプローチがある。ここでは、後者の例であるコミュニケーション状況管理方式の研究を紹介する。

会議システムは、会議における資料の配布、意見の伝達など情報伝達と議事録作成など情報回想の2つの要求を支援する。従来のシステムは、後者に対し議事状況を再利用できるように文字や図で明示的に書き残すものであった。そのため、明示的に表現できない会議の状況、例えば決議に対する納得を現す表情や口調などの曖昧な情報を回想するに苦勞を強いられていた。また、議事録などの作成は個人的視点からの情報利用であるが、このような観点からの情報管理が不足していた。

これらの課題解決には、明示的に表現できない会議の状況を映像や音声をはじめとするマルチメディア情報で表現し、個人的な視点から利用できるコミュニケーション状況管理方式が必要となる。すなわ

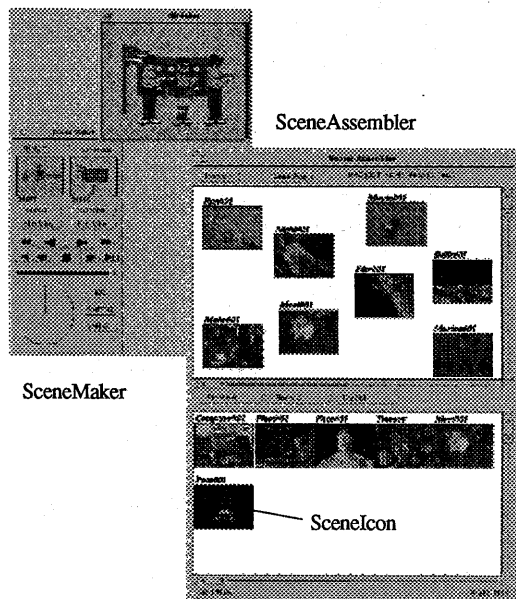


図6 GUI

ち、「状況」を視点とした会議情報の操作と、個人的な会議情報の検索要求に対するユーザモデルの構築を行ない、モデルに基づいた状況情報データベースシステムを実現する。

図7にこのような観点で構成した研究システムのユーザーインターフェイスの示す[8]。

本研究システムでは、会議状況の情報の管理を下記の2つのメカニズムで実現している。

1) 5W1Hユーザモデル： 通常、我々は時、人、場所、理由、方法、いわゆる5W1Hで出来事を表現、整理、理解する事は知られている。そこで、利用者を5W1Hの観点で利用要求を持つユーザーモデルとし、この視点での会議データベース構築を行うシステムとした。従って、本システムでは以下の状況に関するキーでコミュニケーション状況を検索することが可能である。

Who	: 参加者に関わる情報
When	: 会議に関わる時間的な情報
Where	: 会議に関わる場所の情報
What	: 資料に関わる情報
Why, How	: 会議の進行状態に関わる情報 (「問題」、「解決策」など利用者により明言されるキーワード)

2) 個人視点からの情報の管理方法：
 ・参加者が会議開催中に資料にメモなど検索キーを宣言的に付与と、
 ・自動的に検索キーを付与
 による方法を組み合わせた方法としている。

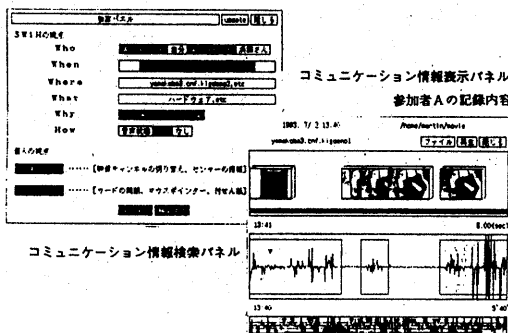


図7 会議システムのインターフェイス

後者は、個人の印象に残り、且つ、比較的、計算機により検知しやすい情報を付与している。具体的には、会議システムの利用資源の変化、例えばウィンドウやマウスの操作の状況をコミュニケーション状況ともに記録・利用する。従って、利用者は利用資源の変化を検索キーとして会議状況を検索することが可能である。

このようにマルチメディア情報を活用したコミュニケーション状況の記録は、従来困難であった曖昧な情報の活用が可能となり会議情報を本質的に活用する可能性を指向したものとなっている。

7. おわりに

このように次世代のメディア技術は今までにないアクティブなエンターテイメントやクリエイティブな生活体験を実現するものと考えられる。また、述べてきたようにマルチメディアによって感性に踏み込む技術、いずれも人間にとっての価値を求める研究となっている。それ故、ここに実現される次世代の技術は音・映像・情報環境の革新を通して真に人間中心の新しい技術への発展が予感される。

[参考文献]

- 1] 吉岡, 他: 「リアルタイム画像生成システムAVIPのシステムアーキテクチャ」 情報処理学会第47回全国大会 (Oct. 1993)
- 2] 藤井, 他: 「リアルタイム画像生成システム用OSの機能と実装」 情報処理学会, システムソフトウェアとOS研究会資料 (Jul. 1993)
- 3] 大谷, 他: 「ラジオシティ法の一並列化手法」 情報処理学会, グラフィックスとCADシンポジウム (1993)
- 4] 原島: 「顔画像の分析符号化方式の動向」 TV学技報, ICS88-17 (Feb. 1988)
- 5] 角, 中川, 他: 「表情・年齢変化の可能な3次元画像合成システム」 電子情報通信学会, HC91-43 (Feb. 1992)
- 6] 谷口, 他: 「マルチメディア指向編集システムの開発」 情報処理学会全国大会 (1993)
- 7] 渡辺, 他: 「マルチメディア分散在席会議システムMERAID」 情報処理学会論文誌, Vol. 32, No. 9 (1992)
- 8] 山口, 他: 「個人的視点からのフォーマル・コミュニケーション状況の統合管理手法」 情報処理学会論文誌, Vol. 34, No. 11 (Nov. 1993)